

**Public Broadcasting Metadata Dictionary Project**  
**Dictionary Review Committee**  
**Report & Preliminary Recommendations – September 2002**

## **Background**

As determined at the first meeting of the Metadata Working Group in April 2002, the Dictionary Review Committee was formed as a subset of the Working Group, and began its work in early June 2002. The Committee is comprised of the following members of the Working Group whose organizations have specific metadata-related projects completed or underway:

- Public Broadcasting Service (PBS) – Marilyn Pierce and Cate Twohill (assisted by Carrie Lowe and numerous departments)
- WGBH – Dave MacCarn (assisted by Thom Shepard)
- KUED Media Solutions – Paul Burrows
- Minnesota Public Radio (MPR) – Alan Baker
- National Public Radio (NPR) – Marty Bloss (assisted by Scott Bridgewater)

## **Goal**

As referenced in the Project's April 26 Facilitator's Memo, the Dictionary Review Committee was formed to:

- Determine what to look for and how to present data
- Make sure all relevant dictionaries are known
- Review and recommend the best metadata dictionary for public broadcasting

Among the points of consensus achieved at the April meeting for the work at hand were these notable ones which have underscored the Dictionary Team's work to date:

- *Agreed* – that in developing a metadata dictionary we should consider what already exists and "take the best and leave the rest."
- *Agreed* – that the public broadcasting metadata dictionary has to reflect the various constituencies who will use it.
- *Agreed* – that we need to map against existing standards.

Given approximately 12 metadata-related projects underway amongst Working Group participants, numerous key initiatives with an asset management component within public broadcasting, and approximately 18 stations launching asset management initiatives, the Working Group ambitiously targeted Labor Day 2002 as the target date for its recommendations of a provisional metadata core dictionary.

## Process

To the extent possible, each Dictionary Team member aggregated and represented its organization's *collective* institutional metadata (not just the metadata represented by each Team member's individual project/department). In some instances and respects, these exchanges of metadata-related information represent landmark levels within public broadcasting of inter-departmental and intra- and inter-organizational coordination and consensus building.

Using an Excel template recommended by Alan Baker/MPR, each team participant "harvested" its organization's metadata fields, which were compiled into an aggregate Excel document, grouped by contributing organization.

At this juncture, emerging observations indicated that roughly 90 percent of the aggregate fields overlapped. To better facilitate continuing data manipulation, data sharing, and data reporting, and to definitively identify metadata fields common to the representative organizations of the Dictionary Team, Paul Burrows/KUED Media Solutions imported the data into a Filemaker database, and built a Grouping Tool within the Filemaker database.

Each Dictionary Team member conformed, as closely as possible, their organization's metadata dictionaries to match the Dublin Core Metadata Element Sets, and, using the Grouping Tool, browsed through the collective metadata description fields to group similar fields. Note that three other proposed Elements are used in the Grouping Tool: Audience, Standard (core curriculum and learning objectives) and Holdings. The task to find commonalties amongst the metadata fields resulted in collapsing the lists from over 2,000 entries to 194 common fields, referenced in this report's companion document, [PB Core Metadata Fields](#).

## Status/Expected Outcomes

As ambitious and comprehensive as the Dictionary Team's work has been over the summer vacation months, the Dictionary Team expects to meet in early October to finalize its initial recommendations. This will allow the Dictionary Team (1) to most thoroughly review and hone the common fields, (2) to construct useful definitions for each metadata field, (3) to develop brief guidelines for the proper usage of a particular metadata field (4) to specify existing controlled vocabularies or encoding schemes where possible, and (5) to determine whether the use case/scenario information provided by the Dictionary Project's User Requirements Committee is adequately reflected in the Dictionary Committee's recommendations. The Committee recognizes that there can be a broad range of metadata descriptors, many of which are only of interest to particular communities and specific users within Public Broadcasting, e.g., programmers, traffic, video sales, and distribution systems. Matching our Use Case discoveries

with the recommended metadata fields should yield a priority core set of descriptors for initial rollouts and review, with supplementary descriptors developing as time progresses and user needs and venues emerge.

The rollout or presentation of the recommended Public Broadcasting Metadata Core is two-fold; the first believed to be within the scope of the Metadata Dictionary project, the second to be determined.

1. BROWSE & REVIEW: Once the Metadata Working Group is satisfied with its recommendations, it is time for the rest of the Public Broadcasting community and other interested parties to browse and review the descriptors. The metadata fields, their definitions, guidelines, suggested encoding schemes and preferred application (based on the priorities set by the Use Case reports) can be made available in a variety of formats, including Adobe Acrobat PDF, Microsoft Word, Microsoft Excel and a run-time, browse-only version of a FileMaker Pro database document (does not require end users to own or launch FileMaker Pro). The purpose of the BROWSE & REVIEW phase is to expose the Public Broadcasting communities to the metadata recommendations, but not to provide end user software solutions for actually describing and indexing assets or implementing an asset management system.

2. INTEROPERABILITY SOLUTIONS: These activities may be part of subsequent phase(s) of the larger Public Broadcasting Dictionary project. Recommended for discussion is the extent to which the project's scope includes the provision of interoperability solutions for end users to actually employ and implement the recommended Core. Interoperability solutions may include the conversion of the Public Broadcasting Core into XML, the merging of the Core into MPEG-7 metadata standards and deployment, or the integration of the Core into existing Digital Asset Management systems, such as Virage, Artesia, Convera, IBM, North Plains, NOLA, etc. A standalone, "shrink-wrapped," and affordable database solution may also be considered, one that almost any station or community could afford and deploy. Maintenance and technical support issues loom large, and discussion is encouraged to determine if and how to address them within the scope and mission of this project.